JOURNAL OF NUMBER THEORY 6, 192–200 (1974)

Random Sieves, II

DAVID HAWKINS

Department of Philosophy, University of Colorado,
1511 University Avenue, Boulder, Colorado 80302

Communicated by W. Schmidt

Received October 15, 1971; revised May 10, 1972

A proof is presented that the random sieve, a stochastic analogue of the sieve of Eratosthenes, generates sequences of numbers approximating the density of primes. The expected number $E_n[h]$ of such numbers less than n satisfies

$$E_n[h]/\pi(n) \to 1$$

and the actual number $h(n)$, on any trial, approximates $\pi(n)$ in the sense of the weak law of large numbers. Some additional results are given.

Let an infinite sequence of natural numbers $2, s_1, s_2, \ldots,$ be produced as follows. From the initial sequence $2, 3, 4, \ldots,$ each $n > 2$ is subject to a process of random deletion, having an independent probability 1/2 of surviving. Let s_1 be the first number not deleted. Then a new wave of independent trials occurs, each remaining $n > s_1$ being removed with probability $1/s_1$. The first number s_2 which survives now initiates a third wave, in which each surviving $n > s_2$ is deleted with probability $1/s_2$. This process is repeated incessantly.

In a previous paper the above "random sieve" [1] was developed as a stochastic analogue of the sieve of Eratosthenes, and also of other deterministic sieves such as the "lucky number" sieve described independently by Erdös and Ulam [2, 3]. The latter sieve generates an infinite sequence of survivors with density $\sim 1/\log n$. Some related topics have been discussed in a series of papers [4–6].

A random sieve has a different character from either of the above, in that it does not generate any specific sequence but defines a probability measure on the set of all possible sequences.

The random sieve is conceived in the spirit of providing definitions of normality for sieve-generated sequences. A meaning of "normal" in this context is analogous to that of normal numbers, first studied by Borel. The first result to be proved below is that the expected value $E_n[h]$ of the number h of survivors less than n is $\sim n/\log n \sim \pi(n)$. The second is that any actual $h(n)$ will almost certainly approximate $n/\log n$, in a sense to be defined.

One would like to obtain from the study of an appropriate probabilistic model certain ideas which would lead to fruitful lines of research in the distribution of the actual primes. The present model seems to offer more toward these ends than does the model proposed by Cramér [7], which simply assigns the probability $1/\log n$ to the survival of each $n > 2$. The random sieve embodies a generative algorithm for the primes, whereas Cramér's model presupposes the prime number theorem. The Cramér model treats the occurrence of "primes" as independent events. The theory of the random sieve, on the other hand, owes its principal difficulty and interest to the fact that this assumption of independence is false. The sieving operation intrinsically implies a kind of "feed back" in which the occurrence of a new "prime" decreases the probability for later occurrences. On the other hand the appearance of "twin primes" is more probable than the product of their separate probabilities.

The model considered below is only the simplest random sieve. One may define a series of increasingly restrictive side-conditions, modifying the sieve in the direction of greater resemblance to the prime-number sieve. Thus it may be stipulated that a survivor s will not remove any number $n < s^2$. But it appears good strategy to explore first the simplest case and its implications.

Let a sequence of probability spaces $\{\mathscr{P}_2, \mathscr{P}_3, \ldots\}$ be defined by

$$\mathscr{P}_n = \langle A_n, \mathfrak{A}_n, p_n \rangle, \tag{1}$$

where $A_n = \{2, 3, 4, \ldots, n\}$. $\mathfrak{A}_n$ is the class of subsets of A_n which include 2, and $p_n(B)$ is the probability that all numbers in $B \in \mathfrak{A}_n$ survive a random sieve, while all numbers in $A_n\bar{B} \in \mathfrak{A}_n$ are eliminated.

The probability measure is defined recursively. Let S_n signify the ultimate survival of n in the random sieve. For each $B_n = \{2, s_1, s_2, \ldots, s_i\}$ in $\mathfrak{A}_n$ there exist two sequences in $\mathfrak{A}_{n+1}$, namely $B_n\bar{S}_{n+1} = \{2, s_1, s_2, \ldots, s_i\}$ and $B_nS_{n+1} = \{2, s_1, s_2, \ldots, s_i, n+1\}$. Given $p(B_n)$ on the space $\mathscr{P}_n$, $p(B_nS_{n+1})$ on the space $\mathscr{P}_{n+1}$ is defined by

$$p(B_nS_{n+1}) = p(B_n)\, p(S_{n+1}/B_n), \tag{2}$$

where

$$p(S_{n+1}/B_n) = X(B_n) = \prod_{m \in B_n} \left(1 - \frac{1}{m}\right) \tag{3}$$

is itself a random variable defined on $\mathscr{P}_n$. By the addition $p(B_nS_{n+1}) + p(B_n\bar{S}_{n+1}) = p(B_n)$ it is shown that $\mathscr{P}_n \subset \mathscr{P}_{n+1}$.

The expected value of X_n,

$$E[X_n] = \sum_{B_n \in \mathfrak{A}_n} p(B_n)\, X(B_n) \tag{4}$$
$$= p(S_{n+1})$$
$$= p(S_{n+r}^{(n)}), \qquad r \geqslant 1,$$

where $S_{n+r}^{(n)}$ signifies the survival of $n+r$ up to sieving by survivors not greater than n.

The k-th moment of X_n

$$E[X_n{}^k] = \sum_{B_n \in \mathfrak{A}_n} p(B_n)\, X^k(B_n)$$
$$= p(S_{n+r_1}^{(n)} S_{n+r_2}^{(n)} \cdots S_{n+r_k}^{(n)}), \tag{5}$$

where the r_i, $1 \leqslant i \leqslant k$ are positive and distinct. $\mathfrak{A}_2$ has only one element, 2, and $p(S_2) = 1, p(S_3/A_2) = p(S_3) = \frac{1}{2}$.

Now let

$$f_n(x) = p(X_n = x), \tag{6}$$

$$F_n(x) = p(X_n \leqslant x). \tag{7}$$

Using the Stieltjes integral,

$$E[X_n] = \int_0^\infty x\,dF_n(x), \tag{8}$$

$$E[X_n{}^k] = \int_0^\infty x^k\,dF_n(x). \tag{9}$$

By direct computation,

$$\begin{array}{ll} & \begin{array}{lll} P & x & B \end{array} \\ \hline f_2(x) = & \begin{cases} 1 & \frac{1}{2} \quad 2 \\ 0 & \text{otherwise}, \end{cases} \\ f_3(x) = & \begin{cases} \frac{1}{2} & \frac{1}{2} \quad 2, \bar{3} \\ \frac{1}{2} & \frac{1}{3} \quad 2, 3 \\ 0 & \text{otherwise}, \end{cases} \\ f_4(x) = & \begin{cases} \frac{1}{4} & \frac{1}{2} \quad 2, \bar{3}, \bar{4} \\ \frac{1}{4} & \frac{3}{8} \quad 2, \bar{3}, 4 \\ \frac{1}{3} & \frac{1}{3} \quad 2, 3, \bar{4} \\ \frac{1}{6} & \frac{1}{4} \quad 2, 3, 4 \\ 0 & \text{otherwise}. \end{cases} \end{array}$$

And for $n \geqslant 2$,

$$f_{n+1}(x) = (1 - x) f_n(x) + x \frac{n+1}{n} f_n\left(x \frac{n+1}{n}\right), \tag{10}$$

which follows immediately from (3). Equation (10) may be rewritten in the form

$$\Delta f_n(x) = x \frac{n+1}{n} f_n\left(x \frac{n+1}{n}\right) - x f_n(x). \tag{11}$$

In what follows use is made of the Laplace transform

$$\varphi_n(\lambda) = \int_0^\infty e^{-\lambda x}\, dF_n(x) = E[e^{-\lambda X_n}],$$

of the derivative operator $D_\lambda = d/d\lambda$, and its integral inverse D_λ^{-1}.

The indicated integrations for transforming (11) are

$$\int_0^\infty \Delta\, dF_n(x)\, e^{-\lambda x} = \Delta \varphi_n(\lambda),$$

$$\int_0^\infty x\, dF_n(x)\, e^{-\lambda x} = -D_\lambda \varphi_n(\lambda),$$

while with $x' = x(n+1)/n$, $e^{-\lambda x} = e^{-\lambda x'} e^{\lambda x'/n+1}$,

$$\int_0^\infty x \frac{n+1}{n}\, dF_n\left(x \frac{n+1}{n}\right) e^{-\lambda x} = \int_0^\infty \sum_{k=0}^\infty \frac{(\lambda x')^k}{(n+1)^k\, k!}\, x'\, dF_n(x')\, e^{-\lambda x}$$

$$= -e^{-\lambda D_\lambda/n+1} D_\lambda \varphi_n(\lambda).$$

Thus (11) transforms into

$$\Delta \varphi_n(\lambda) = (1 - e^{-\lambda D_\lambda/n+1})\, D_\lambda \varphi_n(\lambda) \tag{12}$$

Use may now be made of the fact that the k-th negative derivative of the Laplace transform $\varphi(\lambda)$ of $F(x)$ is equal, at $\lambda = 0$, to the k-th moment of the distribution F:

$$E[X_n{}^k] = (-D_\lambda)^k\, \varphi_n(\lambda)|_{\lambda=0}$$

where k is any integer. For convenience one may write $E[X_n{}^k] = E_n[x^k]$.

The indicated operation gives, from (12),

$$\Delta D_\lambda{}^k \varphi_n(\lambda) = \left\{\left(1 - \frac{1}{n+1}\right)^k e^{-\lambda D_\lambda/n+1} - 1\right\} D_\lambda^{k+1} \varphi_n(\lambda)$$

as can be verified inductively by noticing that

$$D_\lambda e^{-\lambda D_\lambda/n+1} = \left(1 - \frac{1}{n+1}\right) e^{-\lambda D_\lambda/n+1} D_\lambda ,$$

$$D_\lambda^{-1} e^{-\lambda D_\lambda/n+1} = \left(1 - \frac{1}{n+1}\right)^{-1} e^{-\lambda D_\lambda/n+1} D_\lambda^{-1}.$$

At $\lambda = 0$ this gives

$$\Delta E_n[x^k] = \left\{\left(1 - \frac{1}{n+1}\right)^k - 1\right\} E_n[x^{k+1}]. \tag{13}$$

Equation (13) may also be derived (for positive k) by elementary probability consideration, using (4) and (5).

For $k = -1$ (13) says

$$\Delta E_n[x^{-1}] = \frac{1}{n}$$

and with initial condition $E_2[x^{-1}] = 2$

$$E_n[x^{-1}] = \sum_2^{n-1} \frac{1}{m} + 2 = L_{n-1} + 1 \tag{14}$$

$$= \log n + O(1).$$

$E_n[x^{-1}]$ is the expected value of the interval $s - n$, where s is the first survivor greater than n.

In dealing with other values of k in (13) a crucial inequality is that of Schwartz [8]. If u and v are arbitrary random variables defined on the same probability space, $E^2[uv] \leqslant E[u^2]\, E[v^2]$. Putting $u = x^{(k-1)/2}$ and $v = x^{(k+1)/2}$,

$$E_n[x^{k+1}]\, E_n[x^{k-1}] \geqslant E_n{}^2[x^k]. \tag{15}$$

Writing (13) with $k = 1$ in the form

$$\Delta E_n^{-1}[x] = \{(n+1)\, E_n^{-1}[x^2]\, E_n{}^2[x] - E_n[x]\}^{-1},$$

together with $k = 1$ in (15), implies

$$\Delta E_n^{-1}[x] \geqslant \frac{1}{n+1} \tag{16}$$

and

$$E_n^{-1}[x] \geqslant L_n , \tag{17}$$

which with (14) implies

$$E_n[x] = \frac{1}{L_n} + O\left(\frac{1}{L_n^2}\right) = \frac{1}{\log n} + O\left(\frac{1}{\log^2 n}\right). \tag{18}$$

A similar argument leads to

$$E_n[x^2] = \frac{1}{L_n^2} + O\left(\frac{1}{L_n^4}\right). \tag{19}$$

Given these results a more refined argument expresses the moments of the random variable x by asymptotic series involving the negative moments—which are computable in closed form—as coefficients. For the first moment of x^{-1}, write $E_n[x^{-1}] = L_{n-1} + 1 = M_n$, so that

$$\begin{aligned}\mu_{k,n} = E_n[(x^{-1} - M_n)^k] &= E_n[x^{-k}] - kM_nE_n[x^{-k+1}] \\ &+ \cdots + (-1)^k M_n^k\end{aligned} \tag{20}$$

defines the k-th central moment of x^{-1}. Taking first differences of both sides of (20) and using (13), plus the fact that

$$\Delta M_n^k = \frac{k}{n} M_n^{k-1} + \binom{k}{2} \frac{1}{n^2} M_n^{k-2} + \cdots,$$

we find that terms in $(1/n)$ drop out of the equation for $\Delta\mu_{k,n}$ and the result gives

$$\mu_{k,n} \to \mu_k \tag{21}$$

so that all central moments are finite.

Now the identity

$$\begin{aligned}M_n x = \frac{1}{1 + \left(\dfrac{1/x - M_n}{M_n}\right)} &= 1 - \frac{(x^{-1} - M_n)}{M_n} + \frac{(x^{-1} - M_n)^2}{M_n^2} \\ &- \cdots + (-1)^\ell \left\{\frac{x(x^{-1} - M_n)^\ell}{M_n^{\ell-1}}\right\}\end{aligned} \tag{22}$$

gives the corresponding equation in terms of expectation values,

$$E_n[x] = \frac{1}{M_n} + \frac{\mu_{2,n}}{M_n^3} - \frac{\mu_{3,n}}{M_n^4} + \cdots + (-1)^\ell \frac{E_n[x(x^{-1} - M_n)^\ell]}{M_n^\ell}. \tag{23}$$

With respect to the last term the Schwartz inequality,

$$E_n^2[x(x^{-1} - M_n)^\ell] \leqslant E_n[x^2]\, E_n[(x^{-1} - M_n)^{2\ell}]$$

leads, from (19), to

$$E_n[x] = \frac{1}{M_n} + \frac{\mu_{2,n}}{M_n^{\;3}} - \cdots + O\left(\frac{1}{M_n^{\ell+1}}\right), \tag{24}$$

where the constants implied by the order function are at most $\pm\mu_{2\ell}^{1/2}$. A similar argument for x^k results in the asymptotic series

$$E_n[x^k] = \frac{1}{M_n^{\;k}}\binom{k+1}{2}\frac{\mu_{2,n}}{M_n^{k+2}} - \binom{k+2}{3}\frac{\mu_{3,n}}{M_n^{k+3}} + \cdots . \tag{25}$$

Let $h(B_n)$ be the number of survivors $\leqslant n$ in the random sieve. In place of (10) one now has

$$\Delta f_n(x, h) = x\frac{n+1}{n} f_n\left(x\frac{n+1}{n}, h-1\right) - xf_n(x, h). \tag{26}$$

Its two-dimensional Laplace transform,

$$\varphi_n(\lambda, \mu) = E_n[e^{-\lambda x - \mu h}],$$

is

$$\Delta\varphi_n(\lambda, \mu) = \left\{1 - e^{-\mu-(\lambda D_\lambda/n+1)}\right\} D_\lambda\varphi_n(\lambda, \mu),$$

which by repeated partial differentiation, with $\lambda = \mu = 0$, implies

$$\Delta E_n[h] = E_n[x], \tag{28}$$

$$\Delta E_n[xh] = \frac{n}{n+1} E_n[x^2] - \frac{1}{n+1} E_n[x^2h], \tag{29}$$

$$\Delta E_n[x^2h] = \left(\frac{n}{n+1}\right)^2 E_n[x^3] + \frac{2n+1}{(n+1)^2} E_n[x^3h], \tag{30}$$

$$\Delta E_n[h^2] = 2E_n[xh] + E_n[x]. \tag{31}$$

The indicated summations may be performed with the aid of the partial summation formula $\Sigma(U_n\Delta V_n) = U_nV_n - \Sigma(V_{n+1}\Delta U_n)$. This leads to

$$E_n[h] = \frac{n-1}{M_n} + \frac{n-1}{M_n^{\;2}} + O\left(\frac{n-1}{M_n^{\;3}}\right), \tag{32}$$

$$E_n[h^2] = \frac{(n-1)^2}{M_n^{\;2}} + 2\frac{(n-1)}{M_n^{\;3}} + O\left(\frac{(n-1)^2}{M_n^{\;4}}\right), \tag{33}$$

so that the variance of h,

$$V_n[h] = E_n[h^2] - E_n^2[h] = O\left(\frac{n^2}{M_n^4}\right). \tag{34}$$

Let $\theta(n)$ be a slowly increasing function. Specifically, let

$$\theta(n) \begin{cases} \to \infty \\ = o(\log n), \end{cases}$$

so that $\theta(n)\, V_n^{1/2}[h]/E_n[h] = o(1)$; then from Chebychev's inequality

$$P\left(\left|\frac{h(n)}{E_n[h]} - 1\right| < \frac{\theta(n)\, V_n^{1/2}[h]}{E_n[h]}\right) > 1 - \frac{1}{\theta^2(n)}. \tag{35}$$

Thus the difference $|\, h(n)/E_n[h] - 1\,|$ can almost certainly be made as small as we please for n large enough. This theorem holds for a single n and does not imply convergence in probability. Equation (31) may be extended, however, as follows. For definiteness choose $\theta_n = \log^{1/3} n$. Choose an infinite sequence of numbers, $(n_1, n_2, \ldots, n_r, \ldots,)$ so that $n_r = [e^{r^2}]$, where $[t]$ is the greatest integer less than t. Then $\theta^2(n_r) = r^{-4/3}$ and for each r

$$P\left(\left|\frac{h(n_r)}{E_{n_r}[h]} - 1\right| < \frac{1}{r^{2/3}}\right) \geqslant 1 - \frac{1}{r^{4/3}}.$$

But the probability that the inequality fails for *any* of the infinite sequence $(n_1, n_2, \ldots,)$ beyond n_r is less than

$$\sum_r^\infty \frac{1}{r^{4/3}} = o(1).$$

Thus almost certainly

$$h(n_r) \sim E_{n_r}[h] \sim \pi(n_r). \tag{36}$$

Another easy strong law result, from the Chebychev inequality applied to $E_n[x^{-1}]$ and $E_n[x^{-2}]$, is that with probability 1 there is an n_0 such that for all $n > n_0$ there is always a survivor between $n^{2+\epsilon}$ and $(n+1)^{2+\epsilon}$.

The general statistical behavior of the sieve may be summarized by comparison with the Cramér model in which $P(S_n) = 1/\log n$ and the S_n are independent events. For that model the variance $V_n[h]$ is $O(n/\log n)$ but not smaller, in comparison with our $V_n[h] = O(n^2/\log^4 n)$.

On the other hand the product $X_n = \prod_{m\in B_n} (1 - 1/m)$, which for the sieve model has a variance $V_n[x] = O(1/\log^4 n)$, in the Cramér model is a product of independent random variables and has a variance which is not better than $O(1/\log^2 n)$.

From the latter aspect of the random sieve we may conclude that fluctuations tend to be self-correcting; from the former, that the period of this self-correction lengthens exponentially, as is suggested by (32). In a single experimental trial of the sieve the value of X_n is likely to wander in the neighborhood of $E_n[x]$ but after a fluctuation the discrepancy will persist for increasingly long periods, and it is this persistence which is responsible for the large variance of $h(n)$. $h(n)$ would have a much smaller variance if measured from a local average density rather than from $E_n[h]$.

Acknowledgments

The author would like to thank Myron Stein, of the Los Alamos Scientific Laboratory, for the numerical calculations of (13) through $n = 1000$. A draft was read by Robert Bell of Harvey Mudd College, who found minor errors, and by Charles Ryavec of this University, who gave very helpful criticism.

References

1. D. Hawkins, The random sieve, *Math. Mag.* **31** (1957–8), 1–3.
2. S. Ulam et al., On certain sequences of integers defined by sieves, *Math. Mag.* **29** (1956), pp. 117–122.
3. P. Erdös and E. Jabotinsky, On sequences generated by a sieving process, *Nederl. Akad. Wetensch. Indag. Math.* **61** (1958), 115–128.
4. D. Hawkins and W. E. Briggs, The lucky number theorem, *Math. Mag.* **31** (1957), 277–280.
5. M. C. Wunderlich, A general class of sieve generated sequences, *Acta Arith.* **16** (1969), 42–56; contains other references.
6. P. Erdös, Some recent advances and current problems in number theory, *Lectures in Modern Mathematics* (ed. T. L. Saaty), vol. III, p. 213, Wiley, New York, 1965.
7. Harold Cramér, On the order of magnitude of the differences between consecutive prime numbers, *Acta Arith.* **2**, 23–46.
8. William Feller, "Probability Theory and its Applications," vol. II, p. 151, Wiley, New York, 1966.